

# Creating Textual Driver Feedback from Telemetric Data

Daniel Braun Ehud Reiter Advait Siddharthan

Department of Computing Science

University of Aberdeen

{r01db14, e.reiter, advait}@abdn.ac.uk

## Abstract

Usage based car insurances, which use sensors to track driver behaviour, are enjoying growing popularity. Although the data collected by these insurances could provide detailed feedback about the driving style, this information is usually kept away from the driver and is used only to calculate insurance premiums. In this paper, we explored the possibility of providing drivers with textual feedback based on telemetric data in order to improve individual driving, but also general road safety. We report that textual feedback generated through NLG was preferred to non-textual summaries currently popular in the field and specifically was better at giving users a concrete idea of how to adapt their driving.

## 1 Introduction

Although the number of road deaths in the UK is steadily decreasing, 1,713 people died in road accidents in 2013 and 21,657 were seriously injured according to the Department for Transport (2014). Nearly 35% of those who died were under the age of 30. Modern cars are often equipped with numerous driving assistance systems that detect and resolve dangerous situations, but these systems are not available in cheaper and older cars, which are particularly popular among younger drivers. In this group so called “black box” or “telematic” car insurances are becoming more and more popular and insurance companies expect that by 2020 nearly 40% of all car insurances in the UK will be telemetric (Rose, 2013).

Telematic insurances use different sensors installed in the car to track the individual driving style of their customers. Instead of calculating insurance premiums based on statistical risk groups,

insurance companies can use these data to create individual risk profiles and calculate insurance premiums accordingly. This offers drivers who belong to a high-risk group, like young male drivers, the opportunity to save money. Very detailed feedback could be produced from these data which could be able to help drivers to improve their driving and hence road safety. However the feedback insurance companies give to their customers, if they give any feedback at all, is often very sparse: The current state of the art of driver feedback, as used by insurance policies like *AXA Drivesave*<sup>1</sup> and *Aviva Drive*<sup>2</sup>, are scores (e.g. from 0 to 100) in general categories like “pace” and “smoothness” or maps where incidents are marked with pins, as used by *Intelligent Marmalade*<sup>3</sup>. As we show in Section 4, this feedback is not perceived as helpful by drivers.

Drivers who use such an insurance have a particularly high motivation (i.e. money) to change their behaviour. However a system which provides helpful feedback could also be useful for other drivers, especially for example for learners and young drivers. Therefore, in this paper, we explored the possibility of providing drivers with individual textual feedback based on telemetric data, in order to improve road safety. We evaluated the concept of textual driver feedback against the current state of the art feedback mechanisms, to find out if a textual feedback system is perceived as more helpful by drivers.

From an NLG point of view there are two main challenges in creating such a system: Driving one hour can create up to 300,000 data points, which have to be grouped and analysed in a way that allows us to describe important information within

<sup>1</sup><https://play.google.com/store/apps/details?id=com.mydrive.axa.drivesave>

<sup>2</sup><https://play.google.com/store/apps/details?id=com.aviva.ukgi.avivadrive>

<sup>3</sup><https://play.google.com/store/apps/details?id=com.wantstudios.marmalade>

this huge amount of data in a short text. And, like all systems that try to achieve a behaviour change, the texts produced by such a feedback system should take psychological considerations into account, in order to increase the likelihood to achieve a behaviour change. This distinguished our work from NLG systems summarising spatio-temporal data in other domains (Turner et al., 2008; Ponnamperuma et al., 2013).

## 2 Related Work

Although earlier work, like Reiter et al. (2003), has shown that behaviour changes are difficult to achieve, we believe that concrete individual driver feedback, based on telemetric data, could contribute to a more secure driving style.

### 2.1 Psychological Aspects of Behaviour Change

There are many theories about how behaviour changes can be achieved. Fogg (2009), for example, identifies three factors which control human behaviour: motivation, ability, and triggers. A similar point was made by Fishbein (2000), who postulated that “any given behaviour is most likely to occur if one has a strong intention to perform the behaviour, if one has the necessary skills and abilities required to perform the behaviour, and if there are no environmental constraints preventing behavioural performance”. Abraham and Michie (2008) defined 26 “generally-applicable behavior change techniques”, like providing information on consequences and providing general encouragement.

### 2.2 Giving Feedback

There is also a huge amount of literature about how to formulate feedback in order to increase the likelihood of having an impact on the recipient. Three popular advices, which were used in this work, are:

Positive feedback is in general perceived as more accurate and correct than negative feedback (Ilgen et al., 1979). **Starting with positive feedback** therefore gives the feedback source more credibility in general, what has a positive influence of the perception and acceptance of possibly following negative feedback (Steelman and Rutkowski, 2004). This technique is often used in clinical settings as part of the so called “feedback sandwich” (Dohrenwend, 2002).

Hattie and Timperley (2007) pointed out, that “**specific goals** are more effective than general or nonspecific ones” (emphasis added).

Ye and Johnson (1995), Teach and Shortliffe (1987), Weiner (1980) and many others pointed out, that it is crucial for the acceptance of feedback from computer systems, that the **feedback is justified** in a way that allows the user to reconstruct how conclusions were drawn.

### 2.3 Feedback Generation

NLG systems that generate feedback have proven to be helpful in many different areas. Gkatzia et al. (2013) for example showed that an NLG system can provide students with feedback that is perceived as helpful as feedback from lecturers, using reinforcement learning. The SkillSum system (Williams and Reiter, 2008), which generates feedback about basic reading skills and performed significantly better than a comparable system that used canned texts. In the context of citizen science, automatically generated feedback has been shown to improve both skill levels and motivation levels among participants (Blake et al., 2012; van der Wal et al., 2016).

As Eugenio et al. (2005) have shown, aggregation is one important factor that influences the effectiveness of feedback generation systems. This is especially important for the system we present in this paper, since it will deal with a huge amount of data.

Another important task, that is closely related to the aggregation, is the identification of important information which will also be an important part of our system. The approach that we present in Section 3.2.2 and Section 3.2.3 is similar to the work from Gatt et al. (2009) and Hallett et al. (2006)

### 2.4 Automotive Behaviour Change Support Systems

Some projects with focus on ecological driving have already successfully used feedback in order to influence driving behaviour: Like Tulusan et al. (2012), who were able to achieve an improvement in fuel efficiency of more than 3% by providing drivers with numerical feedback that was calculated after each route. Boriboonsomsin et al. (2010), who used a combination of instant and non-instant feedback, achieved an average improvement of 6% on city streets and 1% on highways. And Endres et al. (2010) improved fuel effi-

ciency by using social networks and gamification elements.

There are also systems which use instant feedback, like the CarCoach project from Arroyo et al. (2006). CarCoach uses numerous sensors, like cameras and pressure sensors, to provide immediate feedback on incidents like not looking at the road or being distracted by handling the radio while driving. However, Sharon et al. (2005) showed that negative feedback from the system is easily perceived as frustrating. And there is also always a risk that the feedback itself is a further distraction, when given immediately.

### 3 Methods

#### 3.1 Data Collection

Insurance companies use mainly two different approaches to collect their data: They either use permanently installed sensors, often called “black box”, or smart phone applications. In both cases GPS timestamps and coordinates as well as acceleration data are logged. Although especially smart phone solutions, but to a less extent also black box solutions, raise a lot of questions about data reliability and integrity, as pointed out by Händel et al. (2014) and others, according to Nol (2015) these two approaches have together a worldwide market share of nearly 80% of all telematic insurances.

As our research is focused on data analysis and presentation, rather than the collection, we decided to choose a smart phone based approach, as this method is less intrusive for the car owner and can be used by any driver interested in feedback, without going through an insurance company. The application we used for the data collection was based on previous work by Braun et al. (2011).

The data corpus we used to develop our prototype consisted of about 600 road miles, driven by five different drivers in four different countries. Table 1 shows an example of the data logged by the acceleration sensor, Table 2 shows data logged by the GPS receiver. The acceleration sensor logs the date, the time and the acceleration in  $\frac{m}{s^2}$ . The GPS receiver logs the latitude and longitude coordinates, the accuracy of the localization in meters and the GPS timestamp. Additional information that is needed during the data analysis, like street names, street types and speed limits, are obtained from *OpenStreetMap*. In order to access these data, we used *Nominatim*<sup>4</sup>, to match GPS

<sup>4</sup><https://nominatim.openstreetmap.org>

coordinates to streets in *OpenStreetMap*.

#### 3.2 Data Analysis

In order to provide feedback, we first have to decide which behaviour should be classified as “right” and which as “wrong” and when wrong behaviour is relevant or significant enough to be taken into account for the feedback generation.

##### 3.2.1 Specification of Relevant Behaviour

The most obvious approach would probably be to expect law-abiding behaviour. However it is worth considering different points of view before specifying which behaviour should be regarded as “good” and which should be regarded as “bad”. From the police’s point of view the naive approach of law-abidance may be sufficient, from a driving instructor’s point of view other things are also important, like energy-saving and smoothness. As our research is closely related to telematic insurances, particular attention should be paid to the point of view of insurance companies. Although their exact metrics are secret, we know that they take into account speeding, time of day, day of week, acceleration, braking, elapsed distance, road type and other parameters (cf. Händel et al. (2014) for a more extensive list). On one hand we understandably wanted to stick close to the insurance metrics, on the other hand, from a motivational point of view, it is strongly advised to analyse these parameters critically. It would be, for example, very frustrating for a driver who needs to drive to work at 6 a.m. every weekday, to be told that he should not drive before 9 a.m., because it could increase his insurance premium.

After taking all these different considerations into account, we decided to concentrate on speeding and acceleration and braking behaviour. These are three of the most important parameters for insurance companies, because wrong behaviour in these categories often causes accidents. They are also important for driving instructors. There are, of course, many other important parameters, like distraction and safety distance, which can not be taken into account due to the limitation of the available data.

Speeding, acceleration and braking also have quantitative dimensions, which are very important for feedback generation. While it is reasonable to define driving 30 mph where 20 mph are allowed as wrong behaviour it is arguable if that is the case for driving 21 mph too. In the UK, there is no com-

date	time	x	y	z
08.01.2015	12:07:10.838	1.4939818	2.1068976	9.768343
08.01.2015	12:07:10.858	1.4556746	2.183512	9.730036
08.01.2015	12:07:10.879	1.6472107	2.1452048	9.653421

Table 1: Data logged by the acceleration sensor (in  $\frac{m}{s^2}$ )

lat	lon	accuracy (in m)	timestamp
57.16042614	-2.09462595	10.0	1420718831921
57.1604265	-2.0946818	6.0	1420718832933
57.16042663	-2.0946828	6.0	1420718833934

Table 2: Data logged by the GPS receiver

pulsory law about how to handle these issues and the decision is up to the police officer’s discretion. The Association of Chief Police Officers (2015) suggest a tolerance of 10% of the speed limit + 2mph. Other countries have fixed tolerance, like Germany, with a tolerance of 3%, or no tolerance at all, like Switzerland. Due to the limited accuracy of our measuring method, we decided to adopt a tolerance of 10% of the speed limit, before an incident is classified as speeding. We also decided to ignore violations of the speed limit with a length under 10 meters.

While the quantification of speeding incidents can be derived from laws, the situation is less obvious for inappropriate acceleration or braking. After numerous test, we decided to adopt the guidelines we derived from the *AXA Drivesave* app, which categorises speeding and braking incidents in 4 classes: An acceleration up to  $\pm 2 \frac{m}{s^2}$  is permissible. Non-permissible behaviour is classified in three categories: Acceleration between  $\pm 2 - 3 \frac{m}{s^2}$ ,  $\pm 3 - 4 \frac{m}{s^2}$  and  $> \pm 4 \frac{m}{s^2}$ .

### 3.2.2 Detection of Relevant Behaviour

After finishing a trip, the raw sensor data, obtained by the smart phone application, is parsed for incidents that meet the above described criteria. While acceleration and braking incidents can be detected directly from the sensor data, the recognition of speeding needs further information, namely the speed limit. The prototype we developed uses speed limits provided by the OpenStreetMap project. As the speed limit is not available for all streets in the OpenStreetMap-data, we also implemented a fall-back-mechanism, which sets the speed limit to the general national limit for the road type, for example 60 mph for single carriageways in the UK, if no further information is

provided. Although data from OpenStreetMap has shown to be relatively reliable (Neis et al., 2011) user generated data can always have flaws. But since our analysis focuses on recurring behaviour patterns, rather than single incidents, the impact of single failures is minimized. However, for a commercial system, more reliable data sources could be used.

Each detected incident is stored in a database, as shown in Figure 1. The saved data set contains two timestamps and two GPS coordinate-pairs (start and end), the distance of the incident, the maximum value during the incident (either maximum speed or maximum acceleration) and the average value, as well as a unique ID that links to the street the incident happened on.

Based on these information an importance value is calculated for each incident. The importance of an incident is expressed as a number between 0 and 100 and is based on the type of the incident (speeding incidents are more important than braking incidents, which are more important than acceleration incidents), the distance, the maximum and average value and the type of the road the incident happened on.

### 3.2.3 Aggregation through Clustering

Common feedback systems for drivers, like lane departure warning systems or distance alert systems, give instant feedback about current or even upcoming situations. Our approach however is based on non-instant feedback and aims for a weekly feedback period. The significance of a single incident is therefore considerably lower in our system. As past behaviour can not be changed anyway, we focus on influencing future behaviour. We try to achieve this goal by identifying recurring behaviour patterns in the driving as these patterns

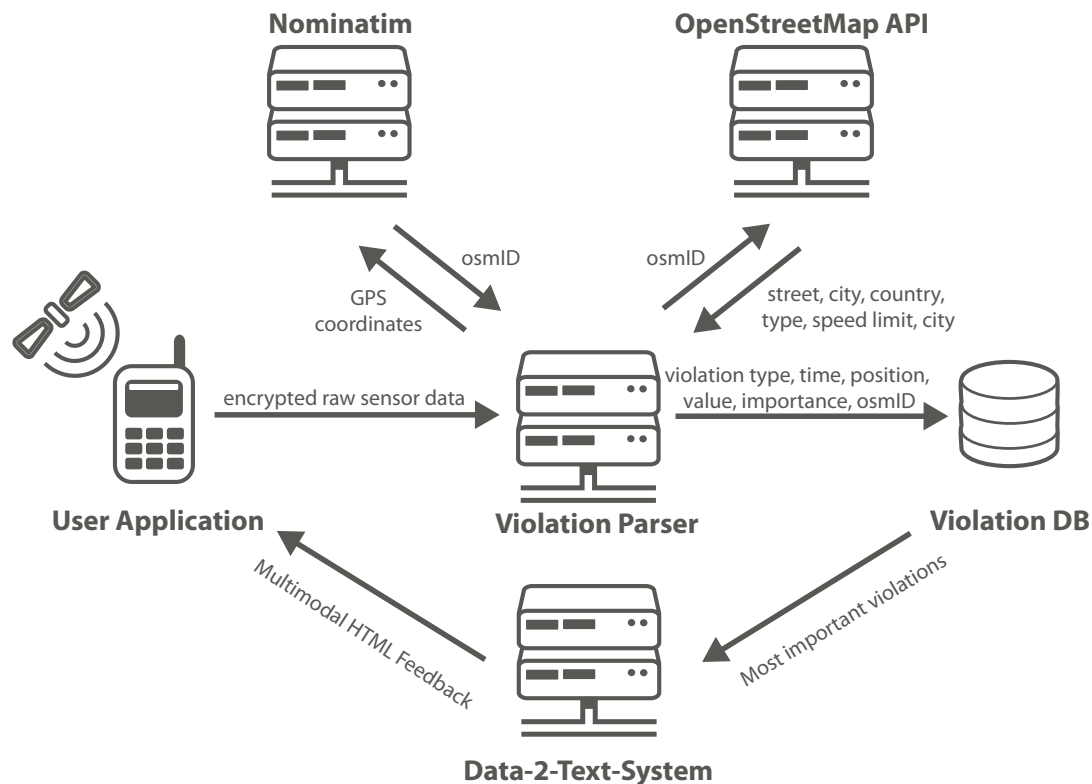


Figure 1: System architecture

are likely to occur again in the future. In this way we hope to not only achieve a change of behaviour in a current situation, but a long-term behaviour change.

Together with domain experts (i.e. driving instructors), we identified features which are suitable to group incidents by, in order to find behaviour patterns: street names, road types, speed limits, time of the day and day of the week. Some of the most common behaviour patterns, according to our domain experts, can be identified by these features. For example the tendency to speed on roads with “extreme” (i.e. very high or very low) speed limits, carelessness on well known routes and dangerous behaviour at certain times (e.g. late in the night or after work).

In order to detect these patterns in the database of all incidents, we use an agglomerative clustering algorithm, where the distance between two incidents is defined by the weighted similarity of all above mentioned features. The algorithm also has a minimal cluster size, which is influenced by the total number of incidents, and a maximum distance, which are used to decide, when to stop the agglomeration and which clusters are irrelevant. In this way we try to balance the interest between greatest possible and tightest possible clus-

ters, since neither very small nor very loose clusters represent significant behaviour patterns.

### 3.3 NLG

The Data-2-Text module of our prototype follows the three-stage pipelined architecture, as described by Reiter (2007), and uses simpleNLG (Gatt and Reiter, 2009) as surface realiser.

#### 3.3.1 Psychological Background

Since we try to achieve a behaviour change, we use different psychological techniques for the verbalisation of feedback, which have been shown to be useful in the literature (cf. Section 2.1) to maximize the likelihood of achieving this goal. This is reflected particularly in the document plan, which follows mainly the three techniques described in Section 2.2. Another psychological aspect was already taken into account during the specification of relevant behaviour. We try to avoid unnecessary frustration by only reporting behaviour that can be easily influenced by the driver, as described in Section 3.2.1.

#### 3.3.2 Document Plan

The high level organisation of the document is based on these ideas. While the number of com-

Summary
Comparison
Map
Single Speeding Incidents
Speeding Clusters
Acceleration Clusters

Table 3: Content order

municated messages differs, depending on the total number of incidents, the order in which the five different message types (in terms of “message types” as used by Reiter and Dale (2000)) are communicated is fixed, as shown in Table 3. The report always starts with a summary, which sums up facts about the reporting period, like the length of the period and the driven distance during this time. The summary is followed, whenever possible, by pointing out a positive development, compared to the last reporting period. This can be very general, if the driver improve broadly, like in Figure 2, “*you reduced the number of speeding incidents per mile by more than 10%*”, or can also be more specific, if the driver did not improve overall, but in one particular aspect, like “*you reduced the number of speeding incidents per mile in residential areas by 20%*”.

After this, a map follows, the main purpose of which is to justify the presented feedback. Each incident is marked with a pin on the map. By clicking on the description of a cluster or a violation type in the text, the map shows only the selected group of incidents and visualizes the frequency of the selected incidents to the user.

Below the map, up to five of the “worst” speeding incidents are reported, described by the amount of speeding and the names of the streets they occurred on. This is only shown if serious speeding, which means exceeding the speed limit by 20 mph or more, happened. Thereupon follows a phrase that specifies how much shorter the braking distance would be, if the driver obeys the speed limit, like “*Going 30 mph slower could shorten your braking distance by 108 yards.*” in the example in Figure 2.

At the end of the report the behaviour patterns, found in form of clusters, are reported. As a short length of the reports is crucial to potential users (c.f. Section 4.6), the number of reported clusters is strictly limited to two of each type, which are selected by their importance. The importance of a

## Driving Report 19 - 25 January

You drove **390 miles in 10 hours and 50 minutes** during the last week. You reduced the number of speeding incidents per mile by **more than 10 %**, well done!

**Five times you drove more than 30 mph too fast:** On Castle Road, on Kirkton Road, on North Deeside Road and twice on A92. Going 30 mph slower could shorten your braking distance by 108 yards. You also **speeded on 175** other occasions, 7 times on **roads with 20 mph speed limit** and 12 times **on weekends on roads with 30 mph speed limit**.

You **accelerated or braked harshly 645** times, mostly on **highways** and on **roads with 20 mph speed limit**.

Figure 2: Feedback type text



Figure 3: Feedback type score

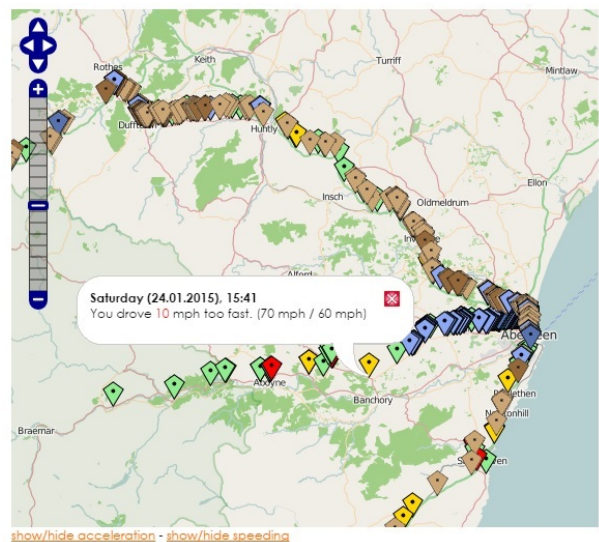


Figure 4: Feedback type map

cluster is a combination of the importance values of the single incidents within the cluster and the size of the cluster.

### 3.3.3 Variation

Due to the fixed structure and the brevity of the text, the space for variation is limited. Nevertheless, as feedback reports will be generated weekly, we added some variation to the text generation. As we expect behaviour changes, there should be a “natural” variation, because of the change of the underlying messages. The most static text, with regard to the underlying data, is the summary at the beginning, therefore there are nine different possibilities how the content of the same message can be realised as text, by changing the order of the sentence, formulations or leaving less important facts, like the driven time, out. In the second part, which starts after the map and consists of two sections, there is also a possible structural variation, as there is either one section about speeding and one about acceleration or one section with single incidents and one with clusters.

## 4 Evaluation

In order to evaluate our approach, we developed a questionnaire to find out how potential users perceive textual feedback, compared to the two state of the art types of feedback, maps and scores.

### 4.1 Data

For this evaluation we used two real datasets recorded in Aberdeen and Aberdeenshire each of which was used twice, once in full length and once by selecting a smaller subset. The feedback that was evaluated by the participants of our study was based on these datasets. These trips were not part of the training dataset we used to develop our prototype.

### 4.2 Questionnaire

We presented feedback reports for four configurations to every participant:

1. low (i.e. short) driven distance, low number of incidents (LL)
2. low driven distance, high number of incidents (LH)
3. high (i.e. long) driven distance, low number of incidents (HL)

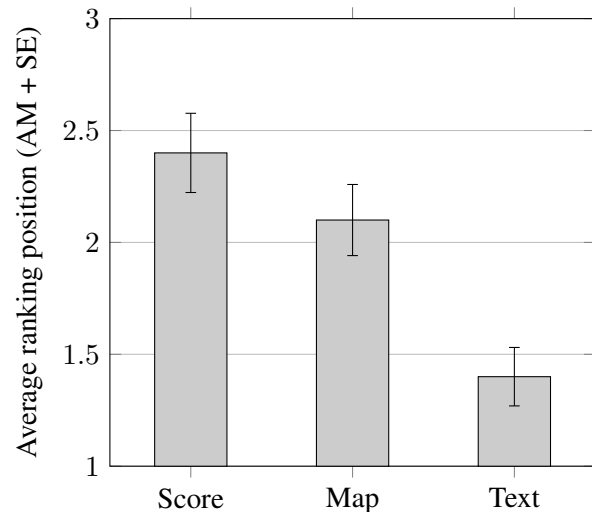


Figure 5: Ranking results

4. high driven distance, high number of incidents (HH)

For each of these four configurations, which were shown in a random order, we presented three types of feedback, which were also shown in a random order: A score, a map and a text. Figures 2, 3 and 4 show the three different types of feedback for the configuration HH. For each type of feedback three statements were given: “The feedback is helpful.”, “The feedback gives me an idea how I could adapt my driving behaviour.” and “The feedback encourages me to change my driving behaviour.”. Participants were asked to indicate how much they agree or disagree with each statement on a Likert scale with seven options. After that, we asked the participants to give a ranking, which type of feedback would be their first, second and third choice, if they had to choose one. We also asked which type(s) of feedback they would choose if they could choose a combination of different types (only one, two or all three). In the end, participants were asked about their attitude towards telematic car insurances in general.

### 4.3 Participants

The survey was completed by 21 participants between the age of 20 and 52. The average age of the participants was 25. About 19% of all participants were female, 81% male. In average the participants had 7 years of driving experience and more than 66% of them drive every day.



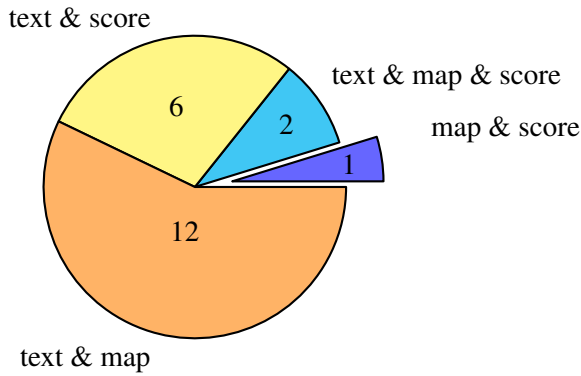


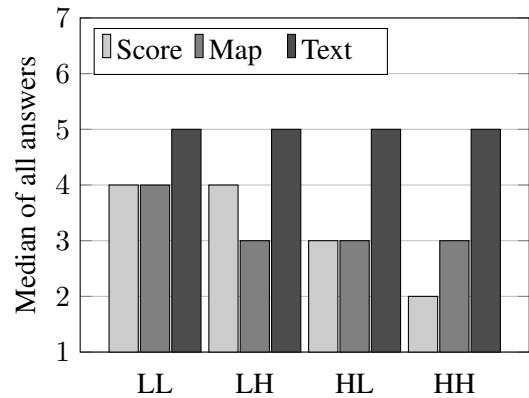
Figure 6: Preferred combination of feedback types

#### 4.4 Basic Findings

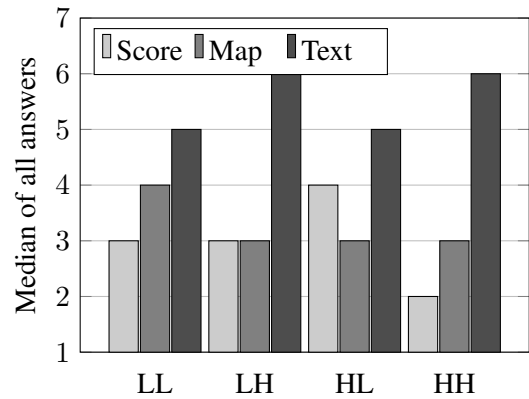
The most basic conclusion that we can draw from the results of this survey is, that our participants preferred the textual feedback over the two other feedback types: 13 participants chose textual feedback as their first preference, 4 the score and 4 the map ( $\chi^2 = 7.722; df = 2; p = 0.02$ ). The average ranking position for the text was 1.4, for the map 2.1 and 2.4 for the score (cf. Figure 5). When asked to choose a combination of feedback types, only one participant chose a combination without textual feedback. The most chosen combination was text and map (12 times). Only two people chose a combination of all three types of feedback (cf. Figure 6).

#### 4.5 Likert Scale Results

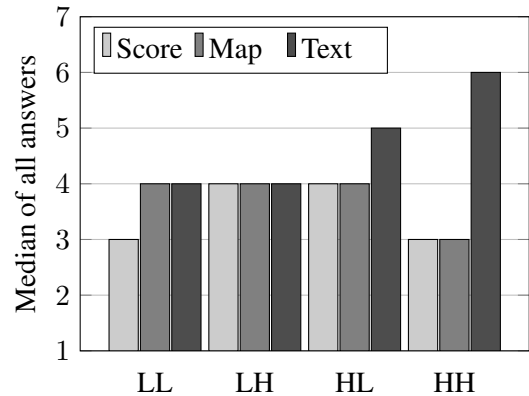
We ran three ANOVA analyses, one with each of the three statements (“The feedback is helpful.”, “The feedback gives me an idea how I could adapt my driving behaviour.” and “The feedback encourages me to change my driving behaviour.”) as dependent variables (Likert scale of 1-7) and feedback type (score, map or text), distance travelled and number of incidents (low/high) as fixed factors and the participant as a random factor. We found an overwhelming main effect of the feedback type ( $p < 0.0001$ ). No other effects or interactions were significant at  $p < 0.05$ . Post hoc analysis by TukeyHSD confirmed that the textual feedback was more helpful, encouraging and provided more ideas than either the map or the score ( $p < 0.0001$  in all cases, except text-map,  $p = 0.0002$ ). Figures 7a to 7c display the differences graphically.



(a) Results for: “The feedback is helpful.”



(b) Results for: “The feedback gives me an idea how I could adapt my driving behaviour.”



(c) Results for: “The feedback encourages me to change my driving behaviour.”

Figure 7: Evaluation of the Likert scale questions (1 = completely disagree, 4 = neutral, 7 = completely agree; LL = distance low & incidents low, LH = distance low & incidents high, HL = distance high & incidents low, HH = distance high & incidents high)



## 4.6 Comments

Six participants used the possibility to give additional comments via a free text field. Three participants said that the length of the text is important and should not be too long. Two participants expressed concerns about the score and that they do not trust the score, because they are not able to reconstruct how it is calculated.

## 4.7 Privacy

Although it was not the focus of our work, we were, of course, aware of the privacy issues that come with a system that tracks locations and analyse behaviour patterns. In our survey, more than 76% of the participants agreed that they would have privacy concerns if they would use a telematic car insurance. Our system itself can run completely autonomously on the phone of the user. That means, in order to guarantee the utmost privacy, no user data will be transmitted. If used in combination with a telematic car insurance, our system does not produce any additional personal data. Instead it processes existing data in a way that, as our evaluation has shown, is more helpful and preferred by users. In this way, the user profits more from his own data and also gets a better understanding of which data is collected.

## 5 Conclusion and Outlook

The results of our evaluation show, that textual driver feedback is perceived as more helpful than the currently used forms of feedback. It also gives drivers a more concrete idea how to adapt their driving. We are confident that textual feedback could not only increase acceptance for automatic generated driver feedback, but could also have a bigger impact on the behaviour than other forms of feedback.

The upcoming EU-legislation “eCall”<sup>5</sup>, which will make telematic sensors mandatory in new cars from April 2018, will lead to a rapid spread of telematic devices in cars within the European Union and will make feedback systems, like the one presented in this paper, even more attractive. Besides the possible applications mentioned above, textual feedback systems could also be used in driving training.

At the moment we are conducting a field study in order to evaluate whether the perceived advan-

<sup>5</sup><http://ec.europa.eu/digital-agenda/ecall-time-saved-lives-saved>

tages of the textual feedback also manifest in a bigger influence on the behaviour of drivers. For this study, we equipped the experimental subjects with smart phone applications, so that each participant will evaluate feedback that is based on his or her own driving and we will be able to analyse if there is a change in behaviour.

## References

- Charles Abraham and Susan Michie. 2008. A taxonomy of behavior change techniques used in interventions. *Health psychology*, 27(3):379.
- Ernesto Arroyo, Shawn Sullivan, and Ted Selker. 2006. Carcoach: A polite and effective driving coach. In *CHI '06 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '06, pages 357–362, New York, NY, USA. ACM.
- Association of Chief Police Officers. 2015. Acpo speed enforcement policy guidelines 2011 - 2015: Joining forces for safer roads. Technical report.
- Steven Blake, Advait Siddharthan, Hien Nguyen, Nirwan Sharma, Anne-Marie Robinson, Elaine O'Mahony, Ben Darvill, Chris Mellish, and Rene Van Der Wal. 2012. Natural language generation for nature conservation: Automating feedback to help volunteers identify bumblebee species. In *COLING*, pages 311–324.
- Kanok Boriboonsomsin, Alexander Vu, and Matthew Barth. 2010. Eco-driving: pilot evaluation of driving behavior changes among us drivers. *University of California Transportation Center*.
- Daniel Braun, Christoph Endres, and Christian Müller. 2011. Determination of mobility context using low-level data. In *Adjunct Proceedings of the 3rd International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI 2011)*, pages 41–42.
- Department for Transport. 2014. Reported road casualties great britain: 2013 annual report. Technical report.
- Anne Dohrenwend. 2002. Serving up the feedback sandwich. *Family practice management*, 9(10):43–50.
- Christoph Endres, Jan Miksatko, and Daniel Braun. 2010. Youldeco-exploiting the power of online social networks for eco-friendly driving. In *Adjunct proceedings of the 2nd International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI 2010)*, page 5.
- Barbara Di Eugenio, Davide Fossati, Dan Yu, Susan Haller, and Michael Glass. 2005. Aggregation improves learning: experiments in natural language

- generation for intelligent tutoring systems. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, pages 50–57. Association for Computational Linguistics.
- Martin Fishbein. 2000. The role of theory in hiv prevention. *AIDS care*, 12(3):273–278.
- Brian J Fogg. 2009. A behavior model for persuasive design. In *Proceedings of the 4th international Conference on Persuasive Technology*, page 40. ACM.
- Albert Gatt and Ehud Reiter. 2009. Simplenlg: A realisation engine for practical applications. In *Proceedings of the 12th European Workshop on Natural Language Generation*, pages 90–93. Association for Computational Linguistics.
- Albert Gatt, Francois Portet, Ehud Reiter, Jim Hunter, Saad Mahamood, Wendy Moncur, and Somayajulu Sripada. 2009. From data to text in the neonatal intensive care unit: Using nlg technology for decision support and information management. *Ai Communications*, 22(3):153–186.
- Dimitra Gkatzia, Helen Hastie, Srinivasan Janarthanam, and Oliver Lemon. 2013. Generating student feedback from time-series data using reinforcement learning. *ENLG 2013*, page 115.
- Catalina Hallett, Richard Power, and Donia Scott. 2006. Summarisation and visualisation of e-health data repositories.
- Peter Händel, Isaac Skog, Johan Wahlström, Farid Bonawiede, Richard Welch, Jens Ohlsson, and Martin Ohlsson. 2014. Insurance telematics: opportunities and challenges with the smartphone solution. *Accepted in IEEE Intell. Transport. Syst. Mag., 2014*.
- John Hattie and Helen Timperley. 2007. The power of feedback. *Review of educational research*, 77(1):81–112.
- Daniel R Ilgen, Cynthia D Fisher, and M Susan Taylor. 1979. Consequences of individual feedback on behavior in organizations. *Journal of applied psychology*, 64(4):349.
- Pascal Neis, Dennis Zielstra, and Alexander Zipf. 2011. The street network evolution of crowdsourced maps: Openstreetmap in germany 2007–2011. *Future Internet*, 4(1):1–21.
- Matthieu Nol. 2015. The role of obd in the 2015 connected car market. *SMis Telematics for Usage-Based Insurance Conference*.
- Kapila Ponnampereuma, Advait Siddharthan, Cheng Zeng, Chris Mellish, and René Van Der Wal. 2013. Tag2blog: Narrative generation from satellite tag data. In *ACL (Conference System Demonstrations)*, pages 169–174.
- Ehud Reiter and Robert Dale. 2000. *Building natural language generation systems*, volume 33. Cambridge university press.
- Ehud Reiter, Roma Robertson, and Liesl M Osman. 2003. Lessons from a failure: Generating tailored smoking cessation letters. *Artificial Intelligence*, 144(1):41–58.
- Ehud Reiter. 2007. An architecture for data-to-text systems. In *Proceedings of the Eleventh European Workshop on Natural Language Generation*, pages 97–104. Association for Computational Linguistics.
- Stuart Rose. 2013. Telematics: How big data is transforming the auto insurance industry. Technical report, SAS.
- Taly Sharon, Ted Selker, Lars Wagner, and Ariel J Frank. 2005. Carcoach: a generalized layered architecture for educational car systems. In *Software Science, Technology and Engineering, 2005. Proceedings. IEEE International Conference on*, pages 13–22. IEEE.
- Lisa A Steelman and Kelly A Rutkowski. 2004. Moderators of employee reactions to negative feedback. *Journal of Managerial Psychology*, 19(1):6–18.
- Randy L Teach and Edward H Shortliffe. 1987. An analysis of physician attitudes regarding computer-based clinical consultation systems. In *Use and impact of computers in clinical medicine*, pages 68–85. Springer.
- Johannes Tulusan, Thorsten Staake, and Elgar Fleisch. 2012. Providing eco-driving feedback to corporate car drivers: what impact does a smartphone application have on their fuel efficiency? In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, pages 212–215. ACM.
- Ross Turner, Somayajulu Sripada, Ehud Reiter, and IanP Davy. 2008. Selecting the content of textual descriptions of geographically located events in spatio-temporal weather data. In Richard Ellis, Tony Allen, and Miltos Petridis, editors, *Applications and Innovations in Intelligent Systems XV*, pages 75–88. Springer London.
- Rene van der Wal, Nirwan Sharma, Anne-Marie Robinson, Chris Mellish, and Advait Siddharthan. 2016. The role of automated feedback in training and retaining conservation volunteers: a case study of bumblebee recording. *To appear in Conservation Biology*.
- JL Weiner. 1980. Blah, a system which explains its reasoning. *Artificial intelligence*, 15(1):19–48.
- Sandra Williams and Ehud Reiter. 2008. Skillsum: basic skills screening with personalised, computer-generated feedback.
- L Richard Ye and Paul E Johnson. 1995. The impact of explanation facilities on user acceptance of expert systems advice. *Mis Quarterly*, pages 157–172.